Jacob, Rebecca (ASRC)

462917

For: Mark Rademacher, AU 3761, 305-0842
US Serial: 09/965681

CP2 3e12

1st Artificial Neural Networks in Medicine and Biology Conference

" Use of a Kohonen Neural Networks to Characterize Respiratory Patients for Medical Intervention"
Kramer, A.A.; Lee, D.; Axelrod, R.C.
Page: 192-196
year-(May) 2000

Use of a Kohonen Neural Network to Characterize Respiratory Patients for Medical Intervention

Andrew A. Kramer, Ph.D.

Future Analytics, Inc

Middleburg, VA 20118 USA

Kramer@futureanalytics.com

Diane Lee, M.S.

MEDai, Inc

Orlando, FL USA

Randy C. Axelrod, M.D.

Sentara Health Services

Norfolk, VA USA

Abstract

Chronic Obstructive Pulmonary Disease (COPD) is one of the leading causes of respiratory hospitalisations in adults in the USA. Prognosis correlates highly with early diagnosis, however the disease may go unnoticed in its early stages. A database of 25,000 individuals with respiratory problems was received for further investigation. The reported rate of COPD in this population was 5.8%, which is fairly low. An unsupervised neural network using the Kohonen architecture was applied to the data in order to cluster patients into groups based on risk factors for COPD. The network consisted of five output neurons. After training characteristics of the groups were examined. Three of the groups consisted of patients with a high percent of risk factors for COPD. Patients in two of those groups were correctly diagnosed as having COPD, but patients in the third group were under-diagnosed for COPD. These patients should be re-examined by a pulmonologist for possible treatment of COPD. Thus Kohonen neural networks may be a useful tool for clustering patients into groups for differential medical intervention.

1 Introduction

Chronic obstructive pulmonary disease (COPD) is a disease category that includes emphysema and chronic bronchitis. These diseases are characterized by obstruction to air flow and frequently coexist.

Emphysema causes irreversible lung damage as the walls between the air sacs within the lungs lose their ability to stretch and recoil. Elasticity of the lung tissue is lost, causing air to be trapped in the air sacs and impairing the exchange of oxygen and carbon dioxide. As a result airflow is obstructed. Symptoms of emphysema include cough, shortness of breath and a limited exercise tolerance. Diagnosis is made by pulmonary function tests, along with the patient's history and physical examination. Chronic bronchitis is due to an inflammation and eventual

scarrii
includ
of bre
M
Amer
$32 b
diseas
COPI
mechi
Tr
interv
exam
more
occun
indivi
popul
indivi
D
incide
repor
COPI
discre
group

2

2.1

Data
their
histoi
incluc
aceut
analy

2.2

2.2.1

Expli
Cate
quam
instai
varia

scarring of the lining of the bronchial tubes. Symptoms of chronic bronchitis include chronic cough, increased mucus, frequent clearing of the throat and shortness of breath.

Mortality and morbidity due to COPD is high. An estimated 16 million Americans suffer from COPD, with an annual cost to the nation of approximately $32 billion[1]. The quality of life for a person suffering from COPD diminishes as the disease progresses. At the onset, there is minimal shortness of breath. People with COPD eventually may require supplemental oxygen and may have to rely on mechanical respiratory assistance.

The prognosis for COPD is enhanced considerably by early diagnosis and intervention. However this requires that the person undergo a lengthy physical examination. Usually patients seek care and definitive diagnosis as a result of one or more serious respiratory episodes, after significant tissue damage has already occurred. It would be desirable if a means could be developed of identifying individuals at high risk for developing COPD. Health data on individuals within a population may help identify the combination of characteristics that suggest an individual is likely to develop COPD.

Data on a subset of patients referred to a large health plan were obtained. The incidence of diagnosed COPD in this group of patients was 5.8%, much lower than reported for adults in the U.S. It is likely that these patients were under-diagnosed for COPD. An analysis was undertaken to determine if patients could be clustered into discrete groups based on their health data. The objective was to isolate one or more groups as candidates for increased medical intervention.

2 Materials and Methods

2.1 Description of Population

Data were obtained from a major health care organization in the U.S. on a portion of their subscriber base. The population consisted of 25,615 individuals who had a history of respiratory problems, including asthma. Information for each individual included demographic data, medical conditions and treatments, detailed pharmaceutical data, and health care costs. In all there were over 200 variables available for analysis.

2.2 Analytical Methods

2.2.1 Data Preprocessing

Exploratory analyses were conducted in order to reduce the size of the input space. Categorical variables with a frequency of less than 1.0% were removed, as were quantitative variables with sparse variation. A correlation matrix was created to detect instances of very high multicolinearity (r > 0.9), and one of the highly correlated variables was removed. The remaining variables were normalized to a range of 0 − 1.

2.2.2 *Neural Network Architecture and Processing*

Since there was no outcome variable and the analysis was investigative in nature an unsupervised neural network design was chosen. Specifically, the self-organizing network based on the model initially formulated by Kohonen[2] and described in Simpson[3] was used. The input vector consisted of 68 variables of diverse types (see 2.1 above). There were five nodes in the output layer of the network. The starting neighbourhood size was four and allowed to decrease to zero. All weights were initially set to 0.5 with a learning rate of 0.4. The stopping criteria were a minimum of 400 epochs or a reduction in the learning rate to 0.001, whichever occurred first.

Network training took place using a random presentation of observations, with the distance metric being Euclidean. When training was stopped a data set consisting of the denormalized input vector and the winning output category for each observation was exported to SAS®, where all statistical analyses were conducted.

3 Results

Table 1 gives the frequencies that observations were placed in the five network categories for a single run. (These results were quite stable over additional runs, with group assignment correlations of 90%.) Each group consisted of between 20-25% of the total with the exception being Group 5 containing only 10%.

Table 1: Number of observations placed in each output category

	Group 1	Group2	Group 3	Group 4	Group 5	Total
N	6309	5500	5452	5740	2614	25615
%	24.6%	21.5%	21.3%	22.4%	10.2%	100.0%

Figure 1 shows the percent of respiratory disorders in each of the five groups. Clearly Groups 4 and 5 do not contain patients with COPD. Groups 1 and 2 have similar percentages of patients with chronic bronchitis and emphysema, but patients in Group 3 were less likely to be diagnosed with COPD.
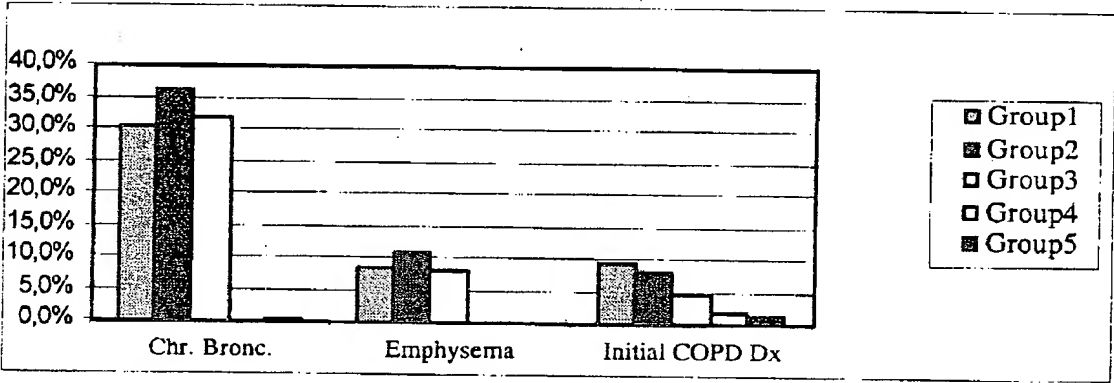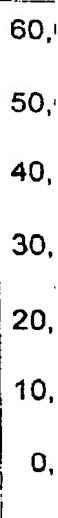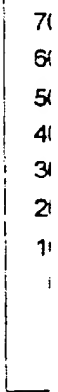

Figure 1: Percent of patients affected with respiratory disorders in the five output groups

. Figure 2 gives the percentage of other chronic conditions for each group. It appears that patients in Groups 1 and 2 have a much higher incidence of other chronic conditions than patients in Group3. Perhaps the lower recorded diagnosis of COPD in Group 3 is due to those patients not being seen as often by physicians.
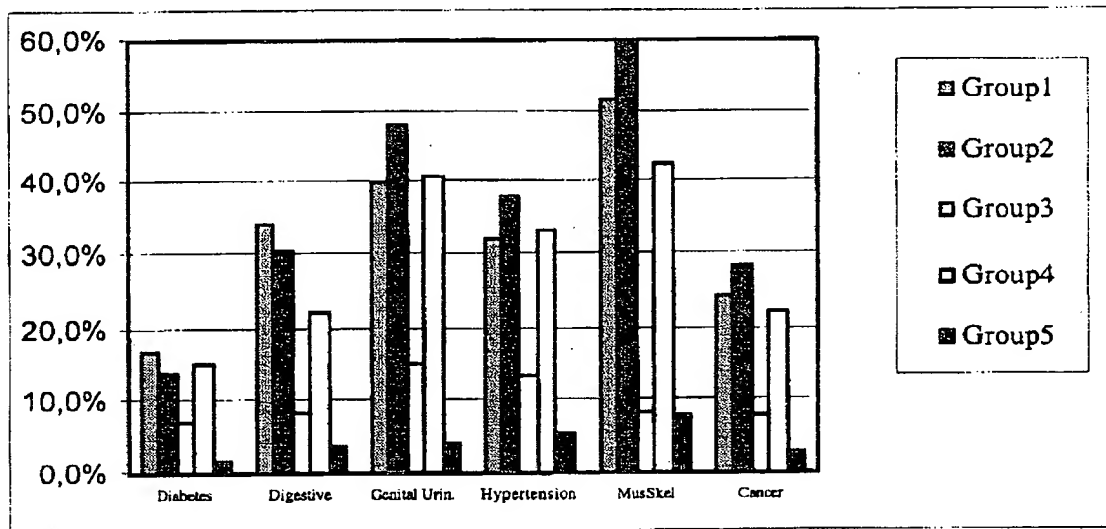

Figure 2: Percentage of patients with other chronic conditions in the five output groups

Figure 3 below shows that patients in Group3 have fewer medical conditions diagnosed and identified drug claims. This results in their costs due to respiratory diseases being a high % of their total medical costs. Group 1 is distinguished from Group 2 in that 100% of the former group had emergency room visits (not shown in Figure 3), as compared to 0% in the latter.
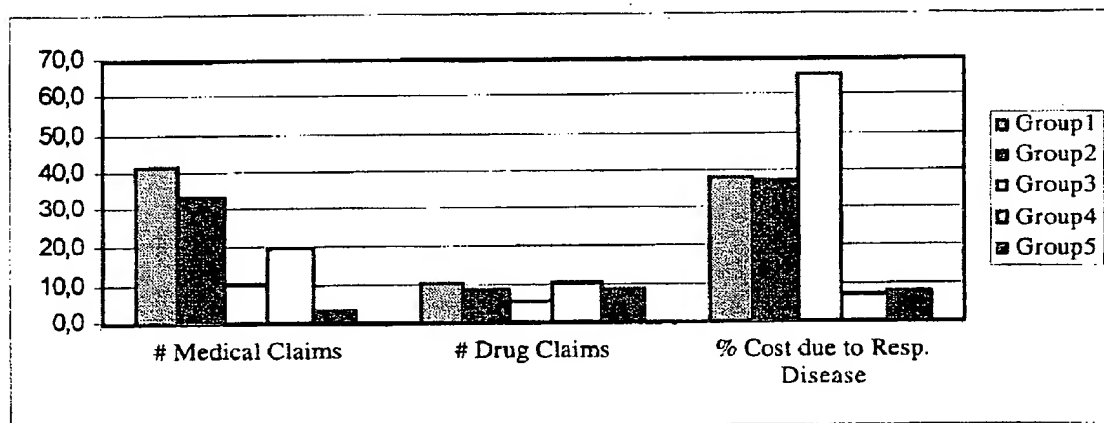

Figure 3: Mean number of other medical conditions, insurance claims, and % of total costs due to respiratory costs

The five groups of patients can be characterized as follows:

1. COPD, multiple medical problems, and at least 1 emergency room visit

2. COPD, multiple medical problems, and no emergency room visits

3. Respiratory disease usually unaccompanied by other conditions

4. No COPD, multiple medical problems

5. No COPD, usually no medical problems

4   Conclusion

Using a Kohonen network on medical data allowed for an informative grouping of patients. From this grouping it was apparent that diagnosis of COPD was highly correlated with a patient having respiratory symptoms accompanied by other medical conditions, and thus being evaluated more frequently by a physician. This was especially true for patients who had at least one emergency room visit. Such groupings allow for those organizations responsible for care delivery to approach these populations with distinct care management strategies. The findings here suggest that patients with existing respiratory disease unaccompanied by other chronic conditions should be evaluated more carefully for a diagnosis of COPD.

References

[1] American Lung Association Publication, "Lung Disease Data 1998-99"., New York, 1999.

[2] Kohonen T. The SOM Methodology. In: Deboeck G. and Kohonen T. (eds) Visual Explorations in Finance with Self-Organizing Maps. Springer-Verlag, Berlin, 1998, pp 159-167.

[3] Simpson, P. Artificial Neural Systems. New York, N.Y.: Pergamon Press, 1990.